Two-Generated Commutative Matrix Subalgebras

Thomas J. Laffey and Susan Lazarus
Department of Mathematics
University College Dublin
Belfield
Dublin 4, Ireland

Submitted by Russell Merris

ABSTRACT

Let F be a field, and let $M_n(F)$ be the algebra of $n \times n$ matrices over F. Let $A, B \in M_n(F)$ with $AB = BA$, and let $\mathscr{A}$ be the algebra generated by A, B over F. A theorem of Gerstenhaber [*Ann. Math.* 73:324–348 (1961)] states that the dimension of $\mathscr{A}$ is at most n. Gerstenhaber's proof uses the methods of algebraic geometry. In this paper, we obtain a purely matrix-theoretic proof of the result. We also examine the case when equality occurs. The case where F is algebraically closed and $\mathscr{A}$ is indecomposable (under similarity) holds the key to the general situation, and in the indecomposable case, we obtain a Cayley-Hamilton-like theorem expressing B^k as a polynomial in $I, B, \dots, B^{k-1}$ with coefficients in $F[A]$, where k denotes the number of blocks in the Jordan form of A. If all Jordan blocks of A have the same size, we obtain a nonderogatory-like condition on B which is equivalent to $\dim_F \mathscr{A} = n$. We also show that in this case $\dim_F \mathscr{A} = n$ is equivalent to the maximality of $\mathscr{A}$ as a commutative subalgebra of $M_n(F)$.

INTRODUCTION

Let F be a field, and let $M_n(F)$ be the algebra of $n \times n$ matrices over F. Let $\mathscr{A}$ be a commutative subalgebra of $M_n(F)$. A result of Schur states that $\dim \mathscr{A} \leqslant [n^2/4]+1$. (For a discussion and recent proof, see Herzer and Huppert [5].) At the other extreme, Courter [1] and the first author [6] have constructed maximal commutative subalgebras $\mathscr{A}$ of dimension much less than n. If $\mathscr{A}$ is generated by a single element A, then it is an immediate consequence of the Cayley-Hamilton theorem that $\dim \mathscr{A} \leqslant n$. Gerstenhaber [3] proved the remarkable fact that $\dim \mathscr{A} \leqslant n$ also if $\mathscr{A}$ is generated by two

655 Avenue of the Americas, New York, NY 10010 0024-3795/91/$3.50

elements. In this paper we obtain a new proof of Gerstenhaber's result. We examine when equality occurs, and we show that if the Jordan blocks of a matrix A all have the same size, a commutative subalgebra of $M_n(F)$ generated by A, B is not maximal if its dimension is less than n.

1. A PROOF OF GERSTENHABER'S THEOREM

In this section we obtain an elementary proof of Gerstenhaber's theorem mentioned in the introduction. From correspondence between the first author and Paul Halmos, we understand that Professor Halmos has constructed a proof along similar lines.

The theorem is proved by first exhibiting a basis for the commutative algebra $\mathscr{A} = \operatorname{alg}\langle I, A, B\rangle$ over an algebraically closed field F, where A and B are nilpotent (Theorem 1.1). Gerstenhaber's result (Theorem 1.2) then follows.

THEOREM 1.1. *Let F be an algebraically closed field, and let $A, B \in M_n(F)$ be nilpotent with $AB = BA$. Suppose A has Jordan canonical form $J_{k_1} \oplus \cdots \oplus J_{k_r}$, where $r \geqslant 1$, $k_1 \geqslant k_2 \geqslant \cdots \geqslant k_r \geqslant 1$ are integers, and J_s denotes the Jordan block of size $s \times s$ with eigenvalue 0. Then there is a basis for $\mathscr{A} = \operatorname{alg}\langle I, A, B\rangle$, the algebra generated by I, A, and B, of the form*

$$\begin{aligned}
\mathscr{S} = \{A^iB^j :\quad & 0 \leqslant i \leqslant k_1 - 1 \quad \text{for } j = 0,\\
& 0 \leqslant i \leqslant k_2' - 1 \quad \text{for } j = 1,\\
& \vdots\\
& 0 \leqslant i \leqslant k_r' - 1 \quad \text{for } j = r-1\},
\end{aligned}$$

where $k_1 \geqslant k_2' \geqslant \cdots \geqslant k_r'$ are integers such that $k_i \geqslant k_i'$ for $i = 2, \ldots, r$.

Proof. Since F is algebraically closed, we may assume A is in Jordan canonical form. So we may write

$$A = J_{k_1} \oplus \cdots \oplus J_{k_r},$$

where the k_i $(i = 1, \ldots, r)$ are as above. The equation $AB = BA$ implies that we can write $B = (B_{ij})$, where B_{ij} is a $k_i \times k_j$ block with

$$J_{k_i} B_{ij} = B_{ij} J_{k_j}. \qquad (*)$$

This equation is studied in detail by Gantmacher in [2], and is shown to force B_{ij} to have the following form:

$$B_{ij} = \left(\overbrace{\;0\;}^{k_j - k_i} \,\middle|\, f_{ij}(J_{k_i}) \right) \qquad \text{if} \quad k_i \leqslant k_j$$

and

$$B_{ij} = \left(\begin{array}{c} f_{ij}(J_{k_j}) \\ \hline 0 \end{array} \right) \begin{array}{l} \\ \}k_i - k_j \end{array} \qquad \text{if} \quad k_i \geqslant k_j,$$

where $f_{ij}(x) \in F[x]$, and the number $k_j - k_i$ $(k_i - k_j)$ above (beside) the 0 is the number of columns (rows) of zeros. Conversely, if B has this block form, then $AB = BA$. It is also shown in [2] that the equations $(*)$ yield Frobenius's formula for the dimension of the centralizer, $\mathscr{C}(A)$, of A:

$$\dim \mathscr{C}(A) = \sum_{i=1}^{r} (2i-1)k_i.$$

We aim now to shown that there exists an *A-characteristic* equation for B, i.e. a polynomial with coefficients in $F[A]$ which B satisfies and which generalizes the concept of the standard characteristic equation. Let z be an indeterminate, and consider the ring $F[z]$. For $B = (B_{ij})$ as above, we define a new matrix $\overline{B}(x) \in M_r(F[z])$ as follows:

$$\overline{B}(z) = (\overline{B}(z)_{ij}),$$

where

$$\overline{B}(z)_{ij} = z^{|k_i - k_j|} f_{ij}(z^2).$$

Consider the specialization obtained by taking $z = Z$, where Z is the $2n \times 2n$ matrix

$$Z = \begin{pmatrix} 0 & I_n \\ A & 0 \end{pmatrix}.$$

Then Z satisfies the matrix equation

$$z^2=\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}.$$

So $\bar{B}(Z)_{ij}=Z^{|k_i-k_j|}f_{ij}(Z^2)=Z^{|k_i-k_j|}f_{ij}(A)$, where $f_{ij}(A)$ is identified with the "scalar" matrix

$$\begin{pmatrix} f_{ij}(A) & 0 \\ 0 & f_{ij}(A) \end{pmatrix}.$$

We will henceforth denote $\bar{B}(Z)$ and $\bar{B}(Z)_{ij}$ by $\bar{B}$ and $\bar{B}_{ij}$ respectively.

Now $\bar{B}$ has entries in the commutative ring $F[Z]$. As the Cayley-Hamilton theorem applies in $F[Z]$, we see that

$$\det(xI-\bar{B})(\bar{B})=0$$

where the coefficients of $\det(xI-\bar{B})$ are elements of $F[Z]$. These coefficients are sums of terms of the form

$$\bar{B}_{i_1 i_2}\bar{B}_{i_2 i_3}\cdots\bar{B}_{i_{m-1} i_m}\bar{B}_{i_m i_1},$$

and so the exponent of Z in the coefficient is

$$\begin{aligned}\gamma&=\left(\sum_{\alpha=1}^{m-1}|k_{i_\alpha}-k_{i_{\alpha+1}}|\right)+|k_{i_m}-k_{i_1}|\\ &\equiv\left(\sum_{\alpha=1}^{m-1}\left(k_{i_\alpha}-k_{i_{\alpha+1}}\right)\right)+k_{i_m}-k_{i_1}\equiv 0 \pmod 2.\end{aligned}$$

Hence the coefficients of $\det(xI-\bar{B})$ are of the form

$$\begin{pmatrix} g(A) & 0 \\ 0 & g(A) \end{pmatrix}$$

where $g(x)\in F[x]$.

Now define a function $f:\mathscr{C}(A)\to M_r(F[Z])$ by $f:B\mapsto\bar{B}$, for B any element of $\mathscr{C}(A)$. $M_r(F[Z])$ is a ring and also a module over the ring

$$\left\{\begin{pmatrix} g(A) & \\ & g(A) \end{pmatrix}:g(x)\in F[x]\right\},$$

and

$$f: g(A) \mapsto \begin{pmatrix} g(A) & \\ & g(A) \end{pmatrix}.$$

Let $\overline{\mathscr{C}}$ denote the image of f.

CLAIM. $\mathscr{C}(A) \cong \overline{C}$; *i.e., f is a monomorphism of rings.*

Proof of claim. Clearly $f(B_1 + B_2) = f(B_1) + f(B_2)$. We now show that $f(B_1 B_2) = f(B_1)f(B_2)$.

To simplify notation, in this proof (only), $gh(X)$ denotes $g(X)h(X)$. Suppose that

$$B_1 = \left(B_{ij}^{(1)}\right), \quad B_2 = \left(B_{ij}^{(2)}\right), \qquad 1 \leqslant i, j \leqslant r,$$

where

$$B_{ij}^{(1)} = \begin{cases} \left(\begin{array}{c|c} \overbrace{0}^{k_j - k_i} & f_{ij}(J_{k_i}) \end{array}\right) & \text{if} \quad k_i \leqslant k_j, \\ \left(\begin{array}{c} f_{ij}(J_{k_j}) \\ \hline 0 \end{array}\right) \}k_i - k_j & \text{if} \quad k_i \geqslant k_j \end{cases}$$

and

$$B_{ij}^{(2)} = \begin{cases} \left(\begin{array}{c|c} \overbrace{0}^{k_j - k_i} & h_{ij}(J_{k_i}) \end{array}\right) & \text{if} \quad k_i \leqslant k_j, \\ \left(\begin{array}{c} h_{ij}(J_{k_j}) \\ \hline 0 \end{array}\right) \}k_i - k_j & \text{if} \quad k_i \geqslant k_j, \end{cases}$$

so $\overline{B}_1 = (\overline{B_{ij}^{(1)}})$ and $\overline{B}_2 = (\overline{B_{ij}^{(2)}})$, where $\overline{B}_{ij}^{(1)} = f_{ij}(A) \cdot Z^{|k_i - k_j|}$ and $\overline{B}_{ij}^{(2)} = h_{ij}(A) \cdot Z^{|k_i - k_j|}$.

Now the (i, j)th block of the product $B_1 B_2$ is

$$\sum_{l=1}^{r} B_{il}^{(1)} B_{lj}^{(2)}.$$

There are four cases to consider:

(1) $i \leqslant l \leqslant j$,
(2) $i \leqslant l \geqslant j$,
(3) $i \geqslant l \leqslant j$,
(4) $i \geqslant l \geqslant j$.

In each case we show the term in $\overline{B_1 B_2}$ arising from $B_{il}^{(1)} B_{lj}^{(2)}$ is the same as that in $\overline{B}_1 \overline{B}_2$.

Case 1. $i \leqslant l \leqslant j$. So $k_i \geqslant k_l \geqslant k_j$ and

$$B_{il}^{(1)} B_{lj}^{(2)} = {}_{k_i - k_l}\left\{\begin{pmatrix} f_{il}(J_{kl}) \\ \hline 0 \end{pmatrix}\begin{pmatrix} h_{lj}(J_{k_j}) \\ \hline 0 \end{pmatrix}\right\}_{k_l - k_j} = \begin{pmatrix} f_{il} h_{lj}(J_{k_j}) \\ \hline 0 \end{pmatrix}\Big\}_{k_i - k_j}.$$

So in $\overline{B_1 B_2}$ this contributes $Z^{|k_i - k_j|} f_{il} h_{lj}(A)$. In $\overline{B}_1 \overline{B}_2$,

$$\begin{aligned} \overline{B}_{il}^{(1)} \overline{B}_{lj}^{(2)} &= Z^{|k_i - k_l|} f_{il}(A) \cdot Z^{|k_l - k_j|} f_{lj}(A) \\ &= Z^{k_i - k_j} f_{il} f_{lj}(A). \end{aligned}$$

So $\overline{B}_{il}^{(1)} \overline{B}_{lj}^{(2)} = \overline{B_{il}^{(1)} B_{lj}^{(2)}}$ in case 1.

Case 2. $i \leqslant l \geqslant j$. So $k_i \geqslant k_l \leqslant k_j$ and

$$B_{il}^{(1)} B_{lj}^{(2)} = {}_{k_i - k_l}\left\{\begin{pmatrix} f_{il}(J_{k_l}) \\ \hline 0 \end{pmatrix}\right.\left(\begin{array}{c|c} \overbrace{0}^{k_j - k_l} & h_{lj}(J_{k_l}) \end{array}\right)$$

$$= \begin{pmatrix} \overbrace{0}^{k_j - k_l} & f_{il} h_{lj}(J_{k_l}) \\ 0 & 0 \end{pmatrix}\Big\}_{k_i - k_l}$$

$$= \begin{cases} \begin{pmatrix} J_{k_j}^{k_j - k_l} f_{il} h_{lj}(J_{k_j}) \\ \hline 0 \end{pmatrix}\Big\}_{k_i - k_j} & \text{if} \quad k_i \geqslant k_j \\ \left(\begin{array}{c|c} \overbrace{0}^{k_j - k_i} & J_{k_j}^{k_i - k_l} f_{il} h_{lj}(J_{k_i}) \end{array}\right) & \text{if} \quad k_i \leqslant k_j. \end{cases}$$

So in $\overline{B_1 B_2}$ we get $Z^{|k_i - k_j|} A^{k* - k_l} f_{il} h_{lj}(A)$, where $k_* = \min\{k_i, k_j\}$. The corresponding term in $\overline{B}_1\overline{B}_2$ is

$$\begin{aligned}\overline{B}_{il}^{(1)}\overline{B}_{lj}^{(2)} &= Z^{k_i-k_l}f_{il}(A)Z^{k_j-k_l}h_{lj}(A)\\ &= Z^{k_i+k_j-2k_l}f_{il}h_{lj}(A)\\ &= Z^{|k_i-k_j|}\cdot Z^{2k_*-2k_l}f_{il}h_{lj}(A)\\ &= Z^{|k_i-k_j|}A^{k_*-k_l}f_{il}h_{lj}(A), \qquad \text{as} \quad Z^2=\begin{pmatrix} A & 0\\ 0 & A\end{pmatrix}.\end{aligned}$$

Hence in case 2, $\overline{B}_{il}^{(1)}\overline{B}_{lj}^{(2)} = \overline{B_{il}^{(1)}B_{lj}^{(2)}}$.

Case 3. $i \geqslant l \leqslant j$. So $k_i \leqslant k_l \geqslant k_i$ and

$$B_{il}B_{lj} = \left(\begin{array}{c|c} \overbrace{\;0\;}^{k_l-k_i} & f_{il}(J_{k_i})\end{array}\right)\left(\begin{array}{c} h_{lj}(J_{k_j})\\ \hline 0\end{array}\right)\}k_l-k_j$$

$$= \begin{pmatrix} \overbrace{\;0\;}^{k_l-k_i} & f_{il}h_{lj}(J_{k_j-k_l+k_i})\\ 0 & 0\end{pmatrix}\}k_i-k_j$$

$$= \begin{cases} \left(\begin{array}{c} J_{k_j}^{k_l-k_i}f_{il}h_{lj}(J_{k_j})\\ \hline 0\end{array}\right)\}k_i-k_j & \text{if} \quad k_i \geqslant k_j,\\ \left(\begin{array}{c|c} \underbrace{\;0\;}_{k_j-k_i} & J_{k_j}^{k_l-k_j}f_{il}h_{lj}(J_{k_i})\end{array}\right) & \text{if} \quad k_i \leqslant k_j.\end{cases}$$

So in $\overline{B_1B_2}$ this contributes $Z^{|k_i-k_j|}A^{k_l-k^*}f_{il}h_{lj}(A)$, where $k^* = \max\{k_i, k_j\}$. In $\overline{B}_1\overline{B}_2$ the corresponding term is

$$\begin{aligned}Z^{k_l-k_i}f_{il}(A)Z^{k_l-k_j}h_{lj}(A) &= Z^{2k_l-k_i-k_j}f_{ij}h_{lj}(A)\\ &= Z^{|k_i-k_j|}Z^{2(k_l-k^*)}f_{il}h_{lj}(A)\\ &= Z^{|k_i-k_j|}A^{k_l-k^*}f_{il}h_{lj}(A).\end{aligned}$$

So in case 3, $\overline{B}_{il}^{(1)}\overline{B}_{lj}^{(2)} = \overline{B_{il}^{(1)}B_{lj}^{(2)}}$.

Case 4. $i \geqslant l \geqslant j$. So $k_i \leqslant k_l \leqslant k_j$ and

$$B_{il}^{(1)} B_{lj}^{(2)} = \left(\overbrace{0}^{k_l - k_i} \middle| f_{il}(J_{k_i}) \right) \left(\overbrace{0}^{k_j - k_l} \middle| h_{lj}(J_{kl}) \right)$$

$$= \left(\overbrace{0}^{k_j - k_i} \middle| f_{il} h_{lj}(J_{k_i}) \right).$$

So in $\overline{B_1 B_2}$ this contributes $Z^{|k_i - k_j|} f_{il} h_{lj}(A)$. In $\overline{B}_1 \overline{B}_2$ the corresponding term is

$$\overline{B}_{il}^{(1)} \overline{B}_{lj}^{(2)} = Z^{k_l - k_i} f_{il}(A) Z^{k_j - k_l} h_{lj}(A)$$

$$= Z^{k_j - k_i} f_{il} h_{lj}(A)$$

$$= Z^{|k_i - k_j|} f_{il} h_{lj}(A).$$

So in all four cases $\overline{B}_{il}^{(1)} \overline{B}_{lj}^{(2)} = \overline{B_{il}^{(1)} B_{lj}^{(2)}}$, which shows that

$$f(B_1 B_2) = f(B_1) f(B_2).$$

Thus f is a ring morphism. Clearly f is one-to-one and onto $\overline{\mathscr{C}}$. This proves the claim. ∎

We can now extend f to $\mathscr{C}(A)[x]$ in the obvious way, and then

$$f : \mathscr{C}(A)[x] \cong \overline{\mathscr{C}}[x].$$

Then $\det(xI - \overline{B})(\overline{B}) = 0$ iff $\det(xI - B)(B) = 0$. Thus B satisfies the polynomial $\det(xI - B)$, which is a monic polynomial of degree r with coefficients in $F[A]$. Thus we can write

$$B^r = f_0(A) + f_1(A)B + f_2(A)B^2 + \cdots + f_{r-1}(A)B^{r-1}$$

for some polynomials $f_i(x) \in F[x]$. Thus a basis for $\mathscr{A} = \mathrm{alg}(I, A, B)$ is contained within the set

$$\{A^i B^j : 0 \leqslant i \leqslant k_t - 1,\ 0 \leqslant j \leqslant r - 1\}.$$

We now claim that, in general, a spanning set for $\mathscr{A}$ is

$$\begin{aligned}\mathscr{S}_1 = \{A^iB^j : &\text{for } j=0,\ 0 \leqslant i \leqslant k_1 - 1,\\ &\text{for } j=1,\ 0 \leqslant i \leqslant k_2 - 1,\\ &\text{for } j=2,\ 0 \leqslant i \leqslant k_3 - 1,\\ &\vdots\\ &\text{for } j=r-1,\ 0 \leqslant i \leqslant k_r - 1\}.\end{aligned}$$

We prove the claim by showing that $A^{k_j}B^{j-1}$ can be written as a linear combination of the elements preceeding it in the list. This is done by induction on $j = 1, \dots, r-1$.

For $j = 1$, $A^{k_1} = 0$, so the claim is trivially true. Suppose $j = \alpha \leqslant r-1$, and that the claim holds for all smaller values of j. Write B in the block form

$$B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix},$$

where

$$B_1 = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1\alpha} \\ B_{21} & B_{22} & \cdots & B_{2\alpha} \\ \vdots & \vdots & & \vdots \\ B_{\alpha 1} & B_{\alpha 2} & \cdots & B_{\alpha\alpha} \end{pmatrix},$$

and write

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

with the same block sizes. So $A_1 = J_{k_1} \oplus \cdots \oplus J_{k_\alpha}$. From our earlier work we know there exist polynomials $f_0(x), \dots, f_{\alpha-1}(x)$ in $F[x]$ such that

$$B_1^\alpha = f_{\alpha-1}(A_1)B_1^{\alpha-1} + \cdots + f_1(A_1)B_1 + f_0(A_1).$$

Now B^{α} has the form

$$\begin{pmatrix} B_1^{\alpha} + D_1 & D_2 \\ D_3 & D_4 \end{pmatrix},$$

where each product summand in D_1, D_2, D_3, D_4 contains at least one of the matrices B_2, B_3, or B_4. These matrices B_i $(i = 2,3,4)$ consist of triangularly striped blocks of the form

$$\left(\begin{array}{ccc|c} 0 & \cdots & 0 & \\ \vdots & & \vdots & f(J_{k_\beta}) \\ 0 & \cdots & 0 & \end{array}\right) \quad \text{and} \quad \begin{pmatrix} & f(J_{k_\beta}) & \\ \hline 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix},$$

where $\beta \geqslant \alpha + 1$ (so $k_\beta \leqslant k_{\alpha+1}$), and thus B_i are annihilated by multiplication by $A^{k_{\alpha+1}}$. Thus

$$\begin{aligned} &A^{k_{\alpha+1}}\left[B^{\alpha} - f_{\alpha-1}(A)B^{\alpha-1} - \cdots - f_1(A)B - f_0(A)\right] \\ &\quad = A^{k_{\alpha+1}}\begin{pmatrix} D_1 & D_2 \\ D_3 & D_4 \end{pmatrix} = 0. \end{aligned}$$

Therefore, $A^{k_{\alpha+1}}B^{\alpha} = g_{\alpha-1}(A)B^{\alpha-1} + \cdots + g_1(A)B + g_0(A)$ for some polynomials $g_i(x) \in F[x]$, $i = 1, \dots, \alpha - 1$, and by our induction hypothesis, the degree of $g_i(x)$ is $\leqslant k_{i+1} - 1$. This proves the claim that $\mathscr{S}_1$ is a spanning set for $\mathscr{A}$.

We now refine the set $\mathscr{S}_1$ to the set $\mathscr{S}$ as follows: for each $j = 0, 1, \dots, r-1$ successively, delete A^iB^j from the list $(i = 1, \dots, k_{i-1})$ if it can be written as a linear combination of the elements proceeding it in the list. This yields a new list:

$$\begin{aligned} &I, A, A^2, \dots, A^{k_1 - 1} \\ &B, AB, A^2B, \dots, A^{k_2' - 1}B \\ &B^2, AB^2, \dots, A^{k_3' - 1}B^2 \qquad\qquad (**) \\ &\vdots \\ &B^{r-1}, AB^{r-1}, \dots, A^{k_r' - 1}B^{r-1}, \end{aligned}$$

with $k_i \geqslant k_i'$ for $i = 2, \dots, r-1$. Now firstly, as $A^{k_i'}B^{i-1}$ $(i = 1, \dots, r-2,$

$k_1' = k_1$) can be expressed as a linear combination of elements preceding it in the list, multiplying this expression by B gives $A^{k_i'}B^i$ as a linear combination of elements preceding it in the list. Hence, $k_i' \geqslant k_{i+1}'$. Secondly, if there were a linear relation

$$\sum_{\substack{i=1,\ldots,m\\ j=1,\ldots,l}} \alpha_{ij} A^i B^j$$

among the elements in the above list, with $\alpha_{ml} \neq 0$, then the "last" element, $A^m B^l$, could be expressed as a linear combination of the preceding elements in the list, which contradicts our construction. This shows the list $(**)$ is a basis for $\mathscr{A}$, and so comprises our set $\mathscr{S}$ with the said properties. This completes the proof of Theorem 1.1. ∎

Now Gerstenhaber's result follows as a consequence:

THEOREM 1.2 (Gerstenhaber [3]). *Let F be a field, and let $A, B \in M_n(F)$ with $AB = BA$. Then the algebra* $\mathrm{alg}\langle A, B\rangle$ *generated by A, B over F has dimension at most n.*

Proof. Let $\mathscr{A}$ be the algebra generated over F by A, B, I. Let $\overline{F}$ be the algebraic closure of F. Since

$$\dim_F \mathscr{A} = \dim_{\overline{F}} \overline{\mathscr{A}},$$

where

$$\overline{\mathscr{A}} = \mathscr{A} \otimes_F \overline{F}$$

is the algebra generated by A, B, I over $\overline{F}$, we assume, without loss of generality, that $F = \overline{F}$, i.e., F is algebraically closed.

Next we observe that if $\mathscr{A}$ is decomposable (under similarity), then we may assume

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix},$$

$$B = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix},$$

where $A_i, B_i \in M_{n_i}(F)$ $(i = 1,2)$ where n_1, n_2 are positive integers with $n = n_1 + n_2$. Then $A_i B_i = B_i A_i$, and using induction we may assume

$$\dim_F \operatorname{alg}_F\langle I, A_i, B_i\rangle \leqslant n_i \qquad (i = 1,2).$$

But $\mathscr{A} \subseteq \mathscr{A}_1 \oplus \mathscr{A}_2$, so the result follows. Hence we assume from now on that A is indecomposable.

Since F is algebraically closed, we may assume A is in Jordan canonical form. If A has at least two distinct eigenvalues, we may write $A = A_1 \oplus A_2$, where $A_i \in M_{n_i}(F)$ $(i = 1,2)$ for some positive integers n_i with $n = n_1 + n_2$ and such that A_1, A_2 have no common eigenvalues. But then the equation $AB = BA$ forces a decomposition $B = B_1 \oplus B_2$ compatible with the decomposition of A. But this implies that $\mathscr{A}$ is decomposable, contrary to assumption. Hence $A = aI + N$ where $a \in F$ and N is nilpotent. Since $\mathscr{A}$ contains I, we may assume $a = 0$. Thus we may write

$$A = J_{k_1} \oplus \cdots \oplus J_{k_r},$$

where $r \geqslant 1$, $k_1 \geqslant k_2 \geqslant \cdots \geqslant k_r \geqslant 1$ are integers with $\Sigma_{i=1}^r k_i = n$. Then Theorem 1.1 shows that there exist integers $k_1' \geqslant \cdots \geqslant k_r' \geqslant 1$ with $k_i' \leqslant k_i$ $(i = 1, \dots, r)$ such that the corresponding set $\mathscr{S}$ forms a basis of $\mathscr{A}$. So $\dim \mathscr{A} = \Sigma_{i=1}^r k_i' \leqslant \Sigma_{i=1}^r k_i = n$. This proves the theorem. ■

REMARK. Note that equality holds (i.e. $\dim \mathscr{A} = n$) iff the elements of $\mathscr{S}_1$ form a basis of $\mathscr{A}$.

2. THE HOMOGENEOUS INDECOMPOSABLE CASE

We continue to assume F is algebraically closed and $\mathscr{A}$ is indecomposable. We now restrict our attention to the "homogeneous" case, that is, when the Jordan blocks of A are all th same size, J_k. We generalize the concept of a nonderogatory block matrix and show that $\dim \mathscr{A}$ is n exactly when B is nonderogatory. Further, we show this property characterizes the 2-generated maximal commutative subalgebras of $M_n(F)$. Here we must assume our base field F has characteristic 0.

We assume $A = J_k \oplus \cdots \oplus J_k$, r blocks. So $n = rk$ and $B = [b_{ij}(J)]$, where b_{ij} is a polynomial with coefficients in F. We have seen that B satisfies the A-polynomial (i.e. a polynomial with coefficients in $F[A]$) $\det_{F[A]}(xI - B)$. We will call this the *A-characteristic polynomial* of B. We define an *A-minimal polynomial* of B to be a monic A-polynomial of least degree which B satisfies. If the A-characteristic equation of B is also an A-minimal polynomial of B, we shall say B is *A-nonderogatory*.

LEMMA 2.1. *If $\mathscr{A} = \operatorname{alg}\langle I, A, B\rangle$ has dimension n over the field F, then*

$$B_0 = \left(b_{ij}(0_k)\right)$$

is A-nonderogatory, where 0_k is the $k \times k$ zero matrix.

Proof. Suppose otherwise, that is, B_0 is A-derogatory. Then B_0 satisfies some monic A-polynomial of degree less than r. So we can write

$$B_0^{r-1} = f_{r-2}(A)B_0^{r-2} + \cdots + f_1(A)B_0 + f_0(A)$$

for some polynomials $f_i(A) \in F[A]$, $i = 1,\dots,r-2$. Then, as there exists a matrix $B_1 \in M_r(F[J])$ such that $B = B_0 + AB_1$,

$$B^{r-1} = (B_0 + AB_1)^{r-1} = B_0^{r-1} + AD$$

for some matrix D, using the binomial expansion and the fact that A and B_0 commute. Hence

$$\begin{aligned} A^{k-1}B^{r-1} &= A^{k-1}B_0^{r-1} \\ &= A^{k-1}\left[f_{r-2}(A)B_0^{r-2} + \cdots + f_1(A)B_0 + f_0(A)\right] \\ &= A^{k-1}\left[f_{r-2}(A)(B-AB_1)^{r-2} + \cdots + f_1(A)(B-AB_1) + f_0(A)\right] \\ &= A^{k-1}\left[f_{r-2}(A)B^{r-2} + \cdots + f_1(A)B + f_0(A)\right], \end{aligned}$$

again using the binomial expansion on each summand $A^{k-1}f_i(A)(B - AB_1)^i$ and the facts that A and B commute and $A^k = 0$. Hence $A^{k-1}B^{r-1}$ is expressed as a linear combination over F of the $A^{k-1}B^i$, for $0 \leqslant i \leqslant r-2$. But this contradicts our assumption that $\dim \operatorname{alg}\langle I, A, B\rangle = n$. Thus B_0 is A-nonderogatory.

This implies that the $r \times r$ matrix $B_0' = [b_{ij}(0)]$ (here 0 denotes the integer 0) is nonderogatory in the usual sense; for any polynomial satisfied by B_0' would also be satisfied by B_0. Thus B_0' is similar to an $r \times r$ companion matrix; i.e., there exists an $r \times r$ nonsingular matrix $T' = [t_{ij}]$ over F, and $a_i \in F$, $i = 0, \dots, r-1$, such that

$$T'^{-1}B_0'T' = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 1 \\ a_0 & a_1 & \cdots & a_{r-2} & a_{r-1} \end{bmatrix}.$$

If we now take $T = [t_{ij}I]$, where I is the $k \times k$ identity matrix, then T is invertible and $T^{-1} = [s_{ij}I]$, where $T'^{-1} = [s_{ij}]$. Then

$$T^{-1}B_0T = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & I \\ a_0I & a_1I & \cdots & a_{r-2}I & a_{r-1}I \end{bmatrix},$$

and T commutes with A. ■

THEOREM 2.2. *Under the above hypotheses,* $\dim_F \operatorname{alg}\langle I, A, B\rangle = n$ *if and only if B is similar over* $GL(r, F[J])$ *to an A-companion matrix; i.e., there exists an invertible matrix P with entries from $F[J]$ such that $P^{-1}BP$ is of the form*

$$\begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & I \\ g_1(A) & g_2(A) & \cdots & g_{r-1}(A) & g_r(A) \end{bmatrix}$$

for some polynomials $g_i(A) \in F[A]$, $i = 1, \dots, r$.

Proof. The "if" direction is obvious. So assume $\dim \operatorname{alg}\langle I, A, B\rangle = n$. By the above discussion, we can assume B is of the form

$$B = \begin{bmatrix} Jf_{11}(J) & I + Jf_{12}(J) & \cdots & Jf_{1r}(J) \\ Jf_{21}(J) & Jf_{22}(J) & \cdots & Jf_{2r}(J) \\ \vdots & \vdots & & \vdots \\ Jf_{r-11}(J) & Jf_{r-12}(J) & \cdots & I + Jf_{r-1r}(J) \\ a_0I + Jf_{r1}(J) & a_1I + Jf_{r2}(J) & \cdots & a_{r-1}I + Jf_{rr}(J) \end{bmatrix}.$$

We are looking for a nonsingular $r\times r$ block matrix $P=(p_{ij}(J))$ and polynomials $g_1, g_2, \dots, g_r$ over F such that

$$P^{-1}BP=\begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & I \\ g_1(J) & g_2(J) & \cdots & g_{r-1}(J) & g_r(J) \end{bmatrix}.$$

Thus we need

$$\begin{bmatrix} p_{11}(J) & p_{12}(J) & \cdots & p_{1r}(J) \\ p_{21}(J) & p_{22}(J) & \cdots & p_{2r}(J) \\ \vdots & \vdots & & \vdots \\ p_{r1}(J) & p_{r2}(J) & \cdots & p_{rr}(J) \end{bmatrix}$$

$$\times\begin{bmatrix} Jf_{11}(J) & I+Jf_{12}(J) & \cdots & Jf_{1r}(J) \\ Jf_{21}(J) & Jf_{22}(J) & \cdots & Jf_{2r}(J) \\ \vdots & \vdots & & \vdots \\ Jf_{r-1\,1}(J) & Jf_{r-1\,2}(J) & \cdots & I+Jf_{r-1\,r}(J) \\ a_0I+Jf_{r1}(J) & a_1I+Jf_{r2}(J) & & a_{r-1}I+Jf_{rr}(J) \end{bmatrix}$$

$$=\begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & I \\ g_1(J) & g_2(J) & \cdots & g_{r-1}(J) & g_r(J) \end{bmatrix}$$

$$\times\begin{bmatrix} p_{11}(J) & p_{12}(J) & \cdots & p_{1r}(J) \\ p_{21}(J) & p_{22}(J) & \cdots & p_{2r}(J) \\ \vdots & \vdots & & \vdots \\ p_{r1}(J) & p_{r2}(J) & \cdots & p_{rr}(J) \end{bmatrix}. \tag{1}$$

Let us take $p_{11}(J)=I$ and $p_{1j}(J)=0$ for $j=2,\dots,r$. Then on working out the first row of the product in (1), we see that necessarily

$$p_{21}(J) = Jf_{11}(J),$$

$$p_{22}(J) = I + Jf_{12}(J),$$

$$p_{23}(J) = Jf_{13}(J),$$

$$\vdots$$

$$p_{2r}(J) = Jf_{1r}(J).$$

This then defines $p_{2j}(J) = \delta_{2j}I + Jf_{ij}(J)$ for $j = 1, \dots, r$, where δ_{ij} is the Kronecker delta. Note that p_{11} and p_{22} are of the form $I + Jh_{ii}(J)$ $(i = 1,2)$, while $p_{ij} = Jh_{ij}(J)$ for $i = 1,2$, $j = 1, \dots, r$, $i \neq j$. Now on working out the second row of the product in (1) we get

$$p_{31}(J) = p_{21}(J)Jf_{11}(J) + p_{22}(J)Jf_{21}(J) + \cdots + p_{2r}(J)[a_0 I + Jf_{r1}(J)],$$

$$p_{32}(J) = p_{21}(J)[I + Jf_{12}(J)] + p_{22}(J)Jf_{22}(J)$$

$$+ \cdots + p_{2r}(J)[a_1 I + Jf_{r2}(J)],$$

$$\vdots$$

$$p_{3r}(J) = p_{21}(J)Jf_{1r}(J) + p_{22}(J)Jf_{2r}(J) + \cdots + p_{2r}(J)[a_{r-1}I + Jf_{rr}(J)],$$

thus defining $p_{3j}(J)$ for $j = 1, \dots, r$. Continuing in this fashion, by working out the lth row of the product in (1), for $l = 1, 2, \dots, r-1$, we get an expression for $p_{l+1}\,j$ (for $j = 1, \dots, r$):

$$p_{l+1j}(J) = \left(\sum_{\beta=1}^{r-1} p_{l\beta}[\delta_{\beta j-1}I + Jf_{\beta j}(J)] \right) + p_{lr}[a_{j-1}I + Jf_{rj}(J)].$$

This defines the matrix P. We claim that for $i=j$, $p_{ij}(J)$ is invertible, and in particular of the form $I+Jh_{ii}(J)$ $(i=1,\dots,r)$, and otherwise $p_{ij}(J)$ is of the form $Jh_{ij}(J)$ $(i,j=1,\dots,r,\ i\neq j)$. We have seen this is true when $i=1,2$. So inductively, assume $i>2$ and the remark is true for $p_{i-1,j}$ for $j=1,\dots,r$. Then

$$p_{ii}(J)=\left(\sum_{\beta=1}^{r-1} p_{i-1\beta}(J)\left[\delta_{\beta i-1}I+Jf_{\beta i}(J)\right]\right)+p_{i-1r}(J)\left[a_{i-1}I+Jf_{ri}(J)\right]$$

$$=p_{i-1i-1}(J)\left[I+Jf_{i-1i}(J)\right]$$

$$+\left(\sum_{\substack{\beta=1\\ \beta\neq i-1}}^{r-1} Jf_{\beta i}(J)\right)+p_{i-1r}(J)\left[a_{i-1}I+Jf_{ri}(J)\right].$$

Now by assumption we have $p_{i-1i-1}(J)$ is of the form $I+Jh_{i-1i-1}(J)$, so $p_{i-1i-1}(J)[I+JF_{i-1i}(J)]$ is of the form $I+Jh_i(J)$ for some polynomial h_i. Also, by our inductive hypothesis $p_{i-1r}(J)=Jh_{i-1r}(J)$, so it is clear that $p_{ii}(J)=I+Jh_{ii}(J)$ for some polynomial h_{ii}.

Similarly, if $i\neq j$,

$$p_{ij}(J)=\sum_{\beta=1}^{r-1} p_{i-1\beta}(J)\left[\delta_{\beta j-1}I+Jf_{\beta j}(J)\right]+p_{i-1r}(J)\left[a_{j-1}I+Jf_{rj}(J)\right]$$

$$=p_{i-1j-1}(J)\left[I+Jf_{j-1j}(J)\right]$$

$$+\sum_{\substack{\beta=1\\ \beta\neq j-1}}^{r-1} Jf_{\beta j}(J)+p_{i-1r}(J)\left[a_{j-1}I+Jf_{rj}(J)\right].$$

By the inductive hypothesis $p_{i-1j-1}(J)=Jh_{i-1j-1}(J)$ and $p_{i-1r}(J)=Jh_{i-1r}(J)$; hence it is clear that $p_{ij}(J)=Jh_{ij}(J)$ for some polynomial h_{ij} $(1\leqslant i,j\leqslant r)$.

Now on working out the rth row of the product in (1), we get

$$\begin{aligned}
&g_1(J)+g_2(J)p_{21}(J)+\cdots+g_r(J)p_{r1}(J)\\
&\quad=p_{r1}(J)Jf_{11}(J)+\cdots+p_{r-1\,r}(J)Jf_{r-1\,1}(J)\\
&\qquad\qquad+p_{rr}(J)[a_0I+Jf_{r1}(J)],\\
&g_2(J)p_{22}(J)+\cdots+g_r(J)p_{r2}(J)\\
&\quad=p_{r1}(J)[I+Jf_{12}(J)]+p_{r2}(J)Jf_{22}(J)\\
&\qquad\qquad+\cdots+p_{rr}(J)[a_1I+Jf_{r2}(J)],\\
&\qquad\vdots\\
&g_2(J)p_{2r}(J)+\cdots+g_r(J)p_{rr}(J)\\
&\quad=p_{r1}(J)Jf_{1r}(J)+\cdots+p_{r\,r-1}(J)[I+Jf_{r-1\,r}(J)]\\
&\qquad\qquad+p_{rr}(J)[a_{r-1}I+Jf_{rr}(J)],
\end{aligned}$$

or equivalently, in matrix form,

$$\begin{bmatrix} g_1(J) & g_2(J) & \cdots & g_r(J)\end{bmatrix}P=\begin{bmatrix} z_1(J) & \cdots & z_r(J)\end{bmatrix}$$

where the $z_i(J)$ are the functions on the right-hand sides of the above equations. Then the polynomial functions $g_i(J)$ are (uniquely) defined precisely if P is invertible as a block matrix. But P is of the form $I+N$, where N is nilpotent by our previous remarks. Thus P is invertible with $P^{-1}=I-N+N^2-\cdots+(-1)^{k-1}N^{k-1}$. This then completes the proof of the theorem. ∎

Note that Lemma 2.1 and Theorem 2.2 together assert the equivalence of the following statements for $A=J_k\oplus\cdots\oplus J_k$ ($n=rk$), $AB=BA$:

(1) $\mathscr{A}=\operatorname{alg}\langle I,A,B\rangle$ has dimension n over F,
(2) B_0 is nonderogatory, and
(3) B is similar over $\mathrm{GL}(r,F[J])$ to an A-companion matrix.

We proceed to show that the above statements are equivalent as well to the statement

(4) $\mathscr{A}$ is a maximal commutative subalgebra of $M_n(F)$.

THEOREM 2.3. *Suppose A is similar over F to $J_k \oplus \cdots \oplus J_k$, r blocks, where J_k is a $k \times k$ Jordan block and B and C are $n \times n$ matrices over the field F such that the matrices A, B, and C commute pairwise. Suppose further $\dim_F \mathscr{A} = n$, where $n = rk$ and $\mathscr{A} = \mathrm{alg}\langle I, A, B\rangle$. Then $C \in \mathscr{A}$; that is, $\mathscr{A}$ is a maximal commutative subalgebra of $M_n(F)$.*

Proof. We have seen that we may assume without loss of generality that $A = J_k \oplus \cdots \oplus J_k$, is in Jordan canonical form.

Then by Theorem 2.2, there exists a block matrix $P \in \mathrm{GL}(r, F[J])$ such that PBP^{-1} is an A-companion matrix. By considering PCP^{-1} in place of C, we may assume without loss of generality that

$$A = J \oplus \cdots \oplus J, \qquad B = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & I \\ f_1(J) & f_2(J) & \cdots & f_{r-1}(J) & f_r(J) \end{bmatrix},$$

and $C = [g_{ij}(J)]$, where $J = J_k$. Then in the usual way we can show that $CB = BC$ implies C is an A-polynomial in B:

There exists a vector $v \in F[J]^r$ such that $\mathrm{span}\{v, Bv, B^2v, \dots, B^{r-1}v\} = F[J]^r$; for instance

$$v = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ I \end{bmatrix}$$

is such a vector. Hence $Cv = \sum_{i=0}^{r-1} \alpha_i(J) B^i v$ for some $\alpha_i(J) \in F[J]$. Now $CB = BC$ implies

$$CB^j v = B^j C v = B^j \sum_{i=0}^{r-1} \alpha_i(J) B^i v$$

$$= \left(\sum_{i=0}^{r-1} \alpha_i(J) B^i \right) (B^j v) \qquad (j = 0, \dots, r-1).$$

Thus

$$\left(C - \sum_{i=0}^{r-1} \alpha_i(J) B^i \right) (B^j v) = 0 \qquad (j = 0, \dots, r-1).$$

But $\{B^j v : j = 0, \dots, r-1\}$ is a spanning set for $F[J]^r$. Hence $C - \sum_{i=0}^{r-1}\alpha_i(J)B^i$ annihilates $F[J]^r$. But $F[J]^r$ is a faithful $F[J]^{r\times r}$-module; so $C - \sum_{i=0}^{r-1}\alpha_i(J)B^i$ is the zero matrix. That is, $C = \sum_{i=0}^{r-1}\alpha_i(J)B^i = \sum_{i=0}^{r-1}\alpha_i(A)B^i$. So $C \in \operatorname{alg}\langle I, A, B\rangle$, which proves the result. ∎

In order to prove a converse of Theorem 2.3, we assume that B_0 is A-derogatory (writing $B = B_0 + AB_1$) and show $\dim_F \mathscr{C}(B) > \dim_F \mathscr{A}$. The result follows from the next two lemmas.

Assuming $A = J\oplus \cdots \oplus J$, r blocks of $J = J_k$ $(rk = n)$, and $B = B_0 + B_1 A$, where B_0 is A-derogatory, choose a basis for $\operatorname{alg}\langle I, A, B\rangle = \mathscr{A}$ of the following form:

$$\begin{aligned}
&I, A, \dots, A^{k-1}\\
&B, AB, \dots, A^{l_1-1}B,\\
&B^2, AB^2, \dots, A^{l_2-1}B^2,\\
&\vdots\\
&B^{r-1}, AB^{r-1}, \dots, A^{l_{r-1}-1}B^{r-1}.
\end{aligned} \tag{2}$$

Thus $\dim_F \mathscr{A} = k + l_1 + l_2 + \cdots + l_{r-1} = l$ say. Now let $\mathscr{B} = \mathscr{C}(B_0) \cap M_r(F[J^{k-1}])$.

LEMMA 2.4. *With notation as above*, $\dim_F \mathscr{B} > r$.

Proof. We can assume without loss that B_0 is in Jordan block form (relative to A); i.e.,

$$B_0 = J_{m_1} \oplus \cdots \oplus J_{m_t},$$

where J_{m_i} is of the form

$$J_{m_i} = \begin{bmatrix} 0 & I & 0 & 0 & \cdots & 0\\ 0 & 0 & I & 0 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & 0 & 0 & I\\ 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}$$

where each entry in J_{m_i} is either the $k\times k$ matrix or the $k \times k$ identity matrix, and where $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_t$ and $m_1 + \cdots + m_t = r$. Also $t > 1$.

Then we know that if $X \in \mathscr{C}(B_0)$ then X is of the form $X = [X_{ij}]_{1\leqslant i,j\leqslant t}$, where X_{ij} is an $m_i \times m_j$ triangularly striped block matrix, i.e., X_{ij} has one of the following forms:

$$\begin{pmatrix} f_{11}(J) & f_{12}(J) & \cdots & f_{1m_i}(J) \\ 0 & f_{11}(J) & \cdots & f_{1m_i-1}(J) \\ & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & f_{11}(J) \end{pmatrix},$$

$$\begin{pmatrix} \overbrace{0 \quad \cdots \quad 0}^{m_j - m_i} & f_{11}(J) & \cdots & f_{1m_i}(J) \\ \vdots & \vdots & \ddots & \vdots \\ 0 \quad \cdots \quad 0 & 0 & \cdots & f_{11}(J) \end{pmatrix},$$

or

$$\begin{array}{l} \begin{pmatrix} f_{11}(J) & \cdots & f_{im_j}(J) \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{11}(J) \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \end{array} \Big\} m_i - m_j$$

for $i = j$, $i > j$, and $i < j$ respectively and $f_{ij}(x) \in F[x]$. If $X \in M_r(F[J^{k-1}])$ as well, we must take the functions f_{ij} above to be scalar multiples of J^{k-1}. Thus,

$$\begin{aligned} \dim \mathscr{B} &= m_1 + m_2 + \cdots + m_t + \sum_{1 \leqslant i \neq j \leqslant t} \min\{m_i, m_j\} \\ &> m_1 + m_2 + \cdots + m_t \qquad (\text{as} \quad t > 1) \\ &= r, \end{aligned}$$

which proves the lemma. ∎

Thus, there exists a set of $r+1$ matrices $X_1, X_2, \dots, X_{r+1}$ in $\mathscr{B}$ which are linearly independent over F. Consider the set $\mathscr{S}$ of matrices

$$\begin{aligned}
&X_1, X_2, \dots, X_{r+1},\\
&I, A, \dots, A^{k-2},\\
&B, AB, \dots, A^{l_1-2}B,\\
&B^2, AB^2, \dots, A^{l_2-2}B^2,\\
&\vdots\\
&B^{r-1}, AB^{r-1}, \dots, A^{l_{r-1}-2}B^{r-1},
\end{aligned}$$

where $X_1, \dots, X_{r+1}$ are the independent matrices in $\mathscr{B}$ from above and the remaining matrices are obtained from the list (2) by deleting the last matrix in each of the r rows. Thus $\mathscr{S}$ consists of $l+1$ matrices.

LEMMA 2.5. *$\mathscr{S}$ is a linearly independent set of matrices in $\mathscr{C} = \mathscr{C}_{M_r(F[J])}(B)$.*

Proof. Suppose first of all that

$$0 = \sum_{i=1}^{r+1} \alpha_{0i} X_i + \sum_{i=0}^{k-2} \alpha_{1i} A^i + \sum_{i=1}^{l_1-2} \alpha_{2i} A^i B + \cdots + \sum_{i=0}^{l_{r-1}-2} \alpha_{ri} A^i B^{r-1}$$

for some α_{0i} $(i = 1, \dots, r+1)$, α_{1i} $(i = 0, \dots, k-2), \dots, \alpha_{ri}$ $(i = 0, \dots, l_{r-1})$. Then by multiplying both sides of the equation by A, and noting that $AX_i = 0$ for $i = 1, \dots, r+1$, we see

$$0 = \sum_{i=0}^{k-2} \alpha_{1i} A^{i+1} + \sum_{i=0}^{l_1-2} \alpha_{2i} A^{i+1} B + \cdots + \sum_{i=0}^{l_{r-1}-2} \alpha_{ri} A^{i+1} B^{r-1}.$$

But as the elements in (2) form a basis for $\mathscr{A}$, the above implies $0 = \alpha_{1i}(i = 0, \dots, k-2) = \alpha_{2i}(i = 0, \dots, l_1 - 2) = \cdots = \alpha_{ri}(i = 0, \dots, l_{r-1} - 2)$. But then

$0 = \sum_{i=1}^{r+1}\alpha_{0i}X_i$, and as the $X_1,\dots,X_{r+1}$ are linearly independent over F, we see also that $\alpha_{0i} = 0$ for $i = 0,\dots,r+1$. Thus the lemma is proved. ∎

Thus we have $\dim \mathscr{C} \geqslant l+1 > \dim \operatorname{alg}\langle I, A, B\rangle$, yielding our desired result:

THEOREM 2.6. *With notation as in Theorem 2.3, if $B = B_0 + B_1 A$ and B_0 is A-derogatory, then* $\operatorname{alg}(I, A, B)$ *is not a maximal commutative subalgebra of* $M_n(F)$.

And, finally restating the equivalences we have established:

THEOREM 2.7. *Suppose $A \in M_n(F)$ is similar over F to $J_k \oplus \cdots \oplus J_k$ (r blocks), and suppose $B \in M_n(F)$ with $AB = BA$. Then $B = (b_{ij}(J_k))_{1 \leqslant i,j \leqslant r} \in M_r(F[J_k])$, where $b_{ij}(x) \in F[x]$. Let $B_0 = (b_{ij}(0)) \in M_r(F)$. Then the following are equivalent for $\mathscr{A} = \operatorname{alg}\langle I, A, B\rangle$:*

(i) *$\mathscr{A}$ has dimension n over F;*
(ii) *B_0 is nonderogatory;*
(iii) *B is similar over* $\mathrm{GL}(r, F[J])$ *to an A-companion matrix;*
(iv) *$\mathscr{A}$ is a maximal commutative subalgebra of $M_n(F)$.*

REMARK. We are grateful to the referee for drawing our attention to the work of Guralnick [4]. In Theorem (5.2) of [4], Guralnick obtains a necessary and sufficient condition for an $r \times r$ matrix B with entries in a (commutative) local ring R to be similar over $\mathrm{GL}(r, R)$ to a direct sum

$$C_1 \oplus \cdots \oplus C_m, \tag{$*$}$$

where $C_1,\dots,C_m$ are companion matrices with monic minimal polynomials $f_1(x),\dots,f_m(x)$, respectively, where $f_1 | f_2 | \cdots | f_m$. The ring $F[J]$ is local with maximal ideal $JF[J]$, so Guralnick's work is applicable.

Regard B as an element of $M_r(F[J])$. If B is similar to a direct sum $C_1 \oplus \cdots \oplus C_m$ as in ($*$), then clearly the following conditions are equivalent:

(i) B is A-nonderogatory;
(ii) $\dim \operatorname{alg}\langle I, A, B\rangle = n$;
(iii) $m = 1$.

In general $B \in M_r(F[J])$ is not similar to a direct sum of companion matrices—or even to a block triangular matrix of the form

$$\begin{bmatrix} C_1 & & * & \\ 0 & C_2 & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & C_m \end{bmatrix}$$

[C_i is in $(*)$]—as can be seen on considering for example

$$\begin{pmatrix} 0 & J \\ J & 0 \end{pmatrix}.$$

One can show however that if $\dim \operatorname{alg}\langle I, A, B\rangle = n$, then B does satisfy the condition $\omega(B) = 0$ in [4], and using [4] one can obtain an alternative proof of Theorem 2.2. Guralnick's work [4] is also helpful in analysing $\mathscr{A} = \operatorname{alg}\langle I, A, B, C\rangle$ where A, B, C are commuting matrices with A homogeneous (in the sense of this section) in the indecomposable case.

3. CONCLUSION

The maximal commutative subalgebras of $M_n(F)$ of greatest dimension, first characterized by Schur [3], require $[n^2/4]$ generators. The maximal commutative subalgebras of $M_n(F)$ of very small dimension constructed by Courter [1] and Laffey [6] also require large numbers of generators. (This can be seen immediately from the fact that they all have nilpotent ideals of codimension one and exponent of nilpotency at most three.) Bounds on the dimension of maximal commutative subalgebras with a given number of generators are difficult to obtain. Gerstenhaber's theorem and the results in this paper essentially deal with the 2-generator case. Results for the 3-generator case and other "small"-generator cases will be presented elsewhere.

Finally, we note that an account of related questions on commuting matrices is given in Olga Taussky's beautiful paper [7].

Note added in Proof. Two recent papers have come to our attention in which new proofs of Gerstenhaber's Theorem are presented.

1. J. Barria and P. R. Halmos. Vector bases for two commuting matrices. *Linear Multilinear Alg*. 27:147–157 (1990).
2. A. R. Wadsworth. The algebra generated by two commuting matrices. *Linear Multilinear Alg*. 27:159–162 (1990).

Both contain proofs of Theorem 1.1 above and our treatment is very similar to that in [1].

REFERENCES

1 R. C. Courter, Maximal commutative subalgebras of K_n at exponent three, *Linear Algebra Appl* 6:1–11 (1973).

2 F. R. Gantmacher, *The Theory of Matrices*, Vol. 1, Chelsea, New York, 1959.

3 M. Gerstenhaber, On dominance and varieties of commuting matrices, *Ann. of Math.* (2) 73:324–348 (1961).

4 R. M. Guralnick, Similarity of Matrices over local rings, *Linear Algebra Appl.* 41:161–174 (1981).

5 A. Herzer and B. Huppert, Ein Satz von I. Schur über vertauschbare Matrizen (A theorem of I. Schur on commuting matrices), *Linear Algebra Appl.* 71:151–158 (1985).

6 T. J. Laffey, The minimal dimension of maximal commutative subalgebras of full matrix algebras, *Linear Algebra Appl.* 71:199–212 (1985).

7 O. Taussky, Commutativity of finite matrices, *Amer. Math. Monthly* 64:239–255 (1957).

Received 4 January 1990; final manuscript accepted 23 July 1990